Documenting 'What Nurses Do' — Moving Beyond Coding and Classification.

Charles N. Mead, M.D., M.S.[a,b] and Suzanne Bakken Henry, R.N., D.N.Sc.[a]
[a]School of Nursing and Graduate Group in Medical Information Science
University of California San Francisco, San Francisco, California
[b]CareCentric Solutions, Duluth, Georgia

ABSTRACT

A variety of strategies for knowledge representation have been applied to the texts from a number of medical domains.. Many of the techniques rely on the well-defined ways in which medical terms are used within a given domain, a phenomenon referred to as 'sublanguage.' Because much of nursing documentation involves the use of 'everyday' language, the viable application of sublanguage-based approaches to knowledge representation of nursing documentation is not a forgone conclusion . We propose an approach utilizing semantic markup of nursing notes as a strategy for determining whether the documentation of 'what nurses do' is a sublanguage Results of an initial feasibility study utilizing the approach are presented.

INTRODUCTION

As the healthcare delivery system continues to evolve from a fee-for-service, sickness-care system based around *documentation of services* to a prospective-pay, health-management system centered around *outcomes and the understanding of process*, the importance of an integrated electronic health record (EHR) continues to increase. Conceptually, the EHR will consist of a single 'patient chart,' comprised of all the data from all patient encounters.[1] In actual fact, the EHR will most likely consist of a number of distinct files containing different types of information and/or different abstractions of the same information. For example, the digital output from a set of intraoperative physiologic signal processors to which a patient was attached during gallbladder surgery 10 years ago would be electronically accessible if needed, but would most likely not be in the same file as the list of the patient's current medications, scheduled clinic appointments, and visit notes taken during a ongoing (i.e. current) episode of outpatient treatment for sinusitis and asthma. Users of the outpatient clinic system might find a composite data element for 'Past Surgery' with a value of 'gallbladder surgery, 10 years ago, intraoperative bronchospasm and moderate hypertension.' Clicking a 'Retrieve Record' button might then direct the system to retrieve the complete intraoperative record from a network file server.

Because not all abstractions of patient data can be anticipated *a priori,* data must be collected in a format that allows unanticipated system composition or decomposition. Most notably, if decision support systems are to provide assistance in the form of reminders, protocol suggestions, or meaningful abstractions of historical information, the knowledge representation strategies used to store the original raw data must be robust with respect to the completeness and correctness of the representation. In particular, the data transformations involved in the selected representation strategies must ensure that potentially valuable information is not lost.

Considerable work has been done in knowledge representation of 'physician-centric' text, i.e. the data associated with the issues most often encountered by physicians. [2, 3] The collected papers from the 1997 IMIA Working Group 6 meeting on Natural Language Processing and Medical Concept Representation survey the wealth of work that has been done in the arena of 'medical' informatics, and clearly demonstrate that considerable progress has been made in formally representing the knowledge involved with anatomic description, medical diagnoses and reasoning, and (to some degree) protocol-driven treatment of patients.[4] Of particular interest is the fact that much of the successful application of robust knowledge formalisms stems from the fact that domain-specific medical terms are highly specialized *and* are used in predictable syntactic and semantic constructs, i.e. medical (physician) documentation defines a 'formal sublanguage.' [5]

Less work has been done to define the knowledge representation issues for nursing data.[4,6] Although there is certainly a significant degree of overlap in the *terminology* used by physicians and non-physicians, there are reasons to believe that medical knowledge representation may differ substantively from nursing knowledge representation. In particular, differences in syntactic and semantic structures, and their associated compositional grammars, may affect the cross-domain viability and applicability of certain

1091—8280/97/$5.00

knowledge representation formalisms. These differences stem from several observations, including the fact that nurses are less concerned with organ-system-based, medical diagnosis-driven *descriptive* knowledge, and much more concerned with patient-centric, *interactive* and *process-driven* knowledge. (The obvious indicator of this difference can be seen by simply comparing the number of nouns, adjectives and verbs in a physician note to those in a nursing note. In addition, much of the terminology used in documenting nursing care involves the use of 'everyday' language used in 'everyday' ways.)

To date, considerable effort has been devoted to developing nursing classification systems which focus on capturing 'nursing interventions,' the presumed essence of 'what nurses do.' However, some have argued that because much of nursing knowledge is holistic and therefore difficult to define precisely, much of what nurses do *that affects patient outcomes* cannot be captured by classification systems.[7] The increasing emphasis of payors on linking patient outcomes to cost of care has focused considerable attention on the development of methodologies that precisely (i.e. quantitatively) separate those patient-care activities which positively influence patient outcomes from those whose effect is less tangibly demonstrable. Given the fact that patient outcomes are at least in part the result of the various patient-care processes delivered by nurses, it is becoming increasingly important to capture *in meaningful knowledge formalisms suitable for system analysis* the salient aspects of 'what nurses do.'

Homecare provides an ideal environment for focusing on understanding nursing documentation since it is a domain where nursing plays the primary patient-care role. The Health Care Finance Administration (HCFA) — primary administrators and payors for homecare for the majority of Americans — has recently announced a standardized set of patient outcomes data (the Outcomes Assessment Information Set, OASIS) that will be used to measure patient outcomes across risk-stratified homecare populations.[16] Plans call for homecare providers to submit OASIS data for comparative outcomes analysis with reimbursement tied to the results of these analyses. Because reimbursement will be based on outcomes, providers will be required to understand patient-care processes which result in less-than-optimal patient outcomes. HCFA's change from a 'regulatory' to a 'consumer' perspective as a basis for reimbursement mirrors similar changes in the private payor (HMO) world. The focus on outcomes and process understanding will require the development of knowledge representation strategies which enable patient-care data to be represented in a form suitable for meaningful machine analysis, abstraction, and/or manipulation.

In this paper, we discuss the issues surrounding the question of whether or not the text describing "what nurses do" may be analyzed as a formal sublanguage. We propose an approach utilizing semantic 'markup' of free-text nursing documentation as an initial methodology for answering this question, present the findings of an initial feasibility study utilizing a non-automated implementation of the methodology, and discuss which knowledge formalisms are best suited to representing nursing domain texts in the event that nursing documentation can be shown to be a formal sublanguage.

Standardized Classification Systems for Nursing

To date, the American Nurses Association Steering Committee on Databases to Support Nursing Practice and the development of the Unified Nursing Language System has "recognized" three classification systems for nursing interventions. [8] The Nursing Interventions Classification was developed by expert consensus at the University of Iowa and consists of a three-tiered taxonomy of nursing interventions organized into 6 domains, 27 classes, and 433 interventions with related activities.[9,10] The Home Health Care Classification developed at Georgetown University, and the Omaha System developed by the VNA of Omaha, were both developed more empirically based on retrospective chart review and/or iterative field test use. Both systems use bi-axial classification systems with intervention schema comprising both *targets of interventions* and *modes of action.*[11, 12] Other significant efforts in developing nursing classification systems have focused on the creation of a more 'atomic-level' set of terms describing patient problems, outcomes and patient-care activities, and on utilizing natural language processing techniques to extend and refine the Nursing Intervention Lexicon and Taxonomy.[13, 14]

Recently, Nielsen and Mortensen described 6 axes for representing nursing interventions within the framework architecture for the International Classification of Nursing Practice (ICNP)[15] The axes 'decompose' a nursing intervention into more granular 'component parts': 1) Action Types 2)

Object Types 3) Types of Approaches 4) Means 5) Anatomical Sites and 6) Time/Place. 54 Action Types are further categorized into 5 main action categories: Observing, Managing, Performing, Caring, and Informing. Object Types include both Nursing Phenomena (e.g., pain management) and other Objects (e.g. foley catheter). It is notable that regardless of the axis, no compositional grammar is specified for combining the terms from the various axes into more complex constructs. The absence of such a grammar limits the ability of information processing systems to either decompose or abstract information encoded using the axes, and in particular prevents the automated generation or representation of complex knowledge constructs.

In our previous work, we argued that the existing intervention classification systems are necessary but not sufficient for representing 'what nurses do,' and demonstrated the loss of potentially significant data through irreversible data transformations or abstractions.[17, 18] We have also provided evidence supporting the need for atomic-level terms and associated composition grammars to support a number of computer-based applications including clinical documentation and decision support systems. As an important next step in a program of research focused on representation of nursing terms for computer-based systems, [6, 17, 18] we propose to address two specific questions:

- Do the phrases used to document 'what nurses do' constitute a formal sublanguage?
- Which knowledge representation formalisms are best able to represent nursing documentation?

Representing Nursing Interventions

We have previously argued the existing coding systems may be formally classified *as classification systems* rather than formal terminologies because of their lack of an underlying ontology, i.e. a set of semantically organized concepts that describe a domain including, if necessary, the compositional grammar required to construct complex concepts. [19] In the absence of a formal ontology, a *formal logic* (i.e. a set of rules for performing inferential reasoning about the concepts in the ontology) cannot be specified.

Formal logics are finding increasing application in medical informatics as useful tools for the manipulation of selected sub-domains of physician data (e.g. pathology and radiology reports, specialty clinic decision support systems, etc.). [20, 21] These systems often utilize knowledge representation and manipulation formalisms such as predicate calculus or, more recently, the logically equivalent graphic notation of Conceptual Graphs. [22] (Conceptual graphs have distinct advantages for use in knowledge engineering in healthcare because their iconographic representation and consequent intuitive mapping to natural language make them more accessible to domain experts.) However, before applying tools such as conceptual graphs to encoding nursing interventions, it is essential that the appropriateness be assessed.

In our previous work, we suggested that nursing intervention concepts have *at least* 5 types of semantic relationship.[18] In common with the ICNP framework are "has recipient" and "has delivery mode (type of action)." We also specified "has initiator," "has provider," and "has response" as additional types of relationship nodes in the proposed schema. Implicit in our approach was the notion that the texts of phrases nurses use to document nursing interventions in the patient record constitute a formal sublanguage. However, we did not explicitly examine the question of whether or not the phrases nurses use to document what they do meet the formal criteria for a sublanguage. This is a significant question since the answer should determine the choice of tools chosen to explore the problem.

Sublanguage Definition (Medical Domain)

Based on a review of the body of medical informatics literature, Johnson recently enumerated a number of significant empirical findings regarding the criteria for formal sublanguages.[23] His findings are summarized as follows:

Formal sublanguages

- are definable via the analysis of a relatively small amount of data, i.e. the number of new semantic types and patterns needed to account for structure decreases rapidly as multiple texts are examined.
- contain relatively few number of semantic types into which words and phrases can be grouped.
- contain syntactic groups of terms which serve as objective tests for the semantic distinctions important to the sublanguage.
- have semantic patterns which are characterized by nested structures.
- contain a small number of semantic patterns which represent the essential information units of the domain.

- allow meaningful domain analysis based on semantic patterns rather than syntax.

Using these guidelines, we investigated the use of non-automated semantic 'markup' of free-text nursing notes as a means for determining if these notes could be characterized as defining a formal sublanguage for a particular nursing domain.

METHODS

Our overall approach to this problem is as follows:

- To divide an existing corpus of text related to nursing interventions into two sets: model building and testing.
- Using the model building data set, to apply both non-automated and (later) computer-based tools to syntactically and/or semantically "mark up" the nursing documentation related to interventions.
- To analyze the universe of the marked up phrases to determine the syntactic structures and semantic types and patterns present in the text in the context of criteria for a formal sublanguage.
- Using the results of the analysis, to apply one or more appropriately selected knowledge representations formalisms (e.g. conceptual graphs) to both data sets/
- To test the acceptability of the formalism(s) through a user survey.

This paper reports our initial feasibility study using non-automated semantic markup to identify semantic types. We randomly selected 100 non-redundant phrases from a data set of nursing documentation abstracted verbatim from the charts of 65 homecare patients (NIH-NR02215). Using selected attributes of the architecture proposed for the ICNP and those suggested in our earlier study, we decomposed each intervention statement and assigned semantic attribute types. Frequencies by type were tabulated.

RESULTS

We successfully decomposed all intervention phrases in our data set, and assigned semantic types (attributes) to various nouns, adjectives, and verbs. The most frequently occurring attributes were Action Types (100%), Object Types (99%), Provider (98%), and Recipient (98%). Means (11%), Anatomical Sites (8%), and Time/Place (5%), which were specified in the ICNP intervention architecture, occurred only rarely in this homecare data set. This would be expected given the nature of the care delivered in this setting compared to a more time-focused and invasive acute-care setting.

DISCUSSION

Johnson enumerated six criteria as necessary and sufficient for a body of terminology and its empirical rules of usage to be termed a 'formal sublanguage.' We examined a test data set of homecare nursing documentation for the existence of the first criteria, i.e. a tractable number of semantic types, and found that the data set satisfied this criteria using a set of semantic types proposed by the ICNP and in our earlier work.

Further analysis must be conducted to determine if the homecare texts in our test data set meet the additional criteria enumerated by Johnson. If the language that nurses use to document homecare nursing interventions can, in fact, be shown to satisfy the formal sublanguage criteria—and if, as expected, that conclusion can be rigorously extended to include documentation of non-homecare nursing intervention documentation—then it is our belief that a considerable amount of the knowledge representation experience, perspectives and tool usage garnered in the past 20-plus years of work on medical knowledge representation can be effectively leveraged into the nursing domain. This would provide fertile ground for the considerable amount of much-needed collaborative work required to move state-of-the-art nursing knowledge representation beyond coding and classification. We believe that such an effort is not only necessary given the changing forces in the healthcare delivery system, but also possible if commonalties in the physician and nursing knowledge representation models can be exploited so that the differences may be studied and understood in depth.

CONCLUSION

Based on an initial study using semantic markup to identify core semantic concepts previously identified as potentially fundamental to knowledge representation of nursing documentation, free-text charting of nursing interventions appears to meet at least one of the criteria necessary to consider the text as a formal sublanguage. We plan to continue our work to critically examine more data in terms of the remaining criteria, as well as to automate the process of markup so that larger volumes of data may be more expeditiously analyzed.

Acknowledgments

This work was supported by NIH-NR03874, Representing Nursing Concepts for Computer-Based Systems, S.B. Henry, Principal Investigator. The data set was provided by NIH-NR02215, Quality of Nursing Care of People with AIDS, W.L. Holzemer, Principal Investigator.

1. Dick RS, Steen EB: The computer-based patient record: an essential technology for health care. Washington, D.C.: National Academy Press, 1991.

2. Chute CG, Atken GE, Ihrke DM: An empirical evaluation of concept capture by clinical classifications. In: Lun KC, Degoulet P, Piemme TE, et al., eds. MedInfo92. Geneva, Switzerland: 1992.

3. Chute CG, Cohn SP, E. CK, Oliver DE, Campbell JR: The content coverage of clinical classifications. Journal of the American Medical Informatics Association 1996; 3(3): 224-233.

4. Campbell J, Carpenter P, Sneiderman C, Cohn S, Chute C, Warren J: Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. Journal of the American Medical Informatics Association 1997;4(3): 238-251.

5. Sager N, Lyman M, Bucknall C, Nhan N, Tick L: Natural language Processing and the Representation of Clinical Data. Journal of the American Medical Informatics Association 1994; 1(2):142-160.

6. Henry SB, Holzemer WL, Reilly CA, Campbell KE: Terms used by nurses to describe patient problems: Can SNOMED III represent nursing concepts in the patient record? Journal of the American Medical Informatics Association 1994; 1(1): 61-74.

7. Benner P: From novice to expert: Power and excellence in nursing practice. Palo Alto, California: Addison-Wesley, 1984.

8. McCormick K, Lang N, Zielstorff R, Milholland DK, Saba V, Jacox A: Toward standard classification schemes for nursing language: Recommendations of the American Nurses Association Steering Committee on Databases to Support Nursing Practice. Journal of the American Medical Informatics Association 1994; 1(6): 421-427.

9. McCloskey JC, Bulechek GM: Nursing interventions classification, 2nd ed. St. Louis: C. V. Mosby, 1996.

10. Iowa Intervention Project: The NIC taxonomy structure. Image: Journal of Nursing Scholarship 1993; 25: 187-192.

11. Martin KS, Scheet NJ: The Omaha System: A community health nursing data management model. In: Lun KC, Degoulet P, Piemme TE, et al., eds. MedInfo92, Geneva, Switzerland 1992.

12. Saba VK, Zuckerman AE: A new home health classification method. Caring Magazine 1992; 11(9): 27-34.

13. Ozbolt JG: From minimum data to maximum impact: using clinical data to strengthen patient care. Advanced Practice Nursing Quarterly 1996; 1(4): 62-69.

14. Grobe SJ: The Nursing Intervention Lexicon and Taxonomy: Implications for representing nursing care data in automated records. Holistic Nursing Practice 1996; 11(1): 48-63

15. Nielsen GH, Mortensen RA: The architecture for an International Classification of Nursing Practice (ICNP). International Nursing Review 1996; 43(6): 175-182.

16. Shaughnessy P: Outcome-Based Quality Improvement: National Association of Home Care, 1995.

17. Henry SB, Mead CN: Standardized nursing classification systems: Necessary, but not sufficient for representing what nurses do. In: Cimino JJ, ed. Fall Symposium of the American Medical Informatics Association, Washington, DC, 1996.

18. Henry SB, Mead CN: Nursing classification systems: Necessary but not sufficient for representing what nurses do for inclusion in computer-based patient record systems. Journal of the American Medical Informatics Association 1997; 4(3): 222-232.

19. Ingenerf J: Taxonomic vocabularies in medicine: the intention of usage determines different established structures. In: Groenes RA, Peterson HE, Protti DJ, eds. MedInfo95, Vancouver, British Columbia, 1995.

20. Campbell KE, Musen M: Representation of clinical data using SNOMED III and conceptual graphs. In: Frisse M, ed. Symposium on Computer Applications in Medical Care, Baltimore, MD, 1992.

21. Bernauer J: Conceptual graphs as an operational model for descriptive findings. In: Frisse M, ed. 16th Annual Symposium for Computer Applications in Medical Care, Baltimore, MD, 1992.

22. Sowa J: Conceptual structures. Reading, Massachusetts: Addison Wesley, 1984.

23. Johnson SB: Conceptual graph grammar - A simple formalism for sublanguage. International Medical Informatics Association Working Group 6 Conference on Natural Language and Medical Concept Representation, Jacksonville, FL 1997.